Ion Torrent semiconductor sequencing allows rapid, low cost sequencing of the human exome

David Jenkins
Bioinformatics Engineer
EdgeBio

Contract Research Division

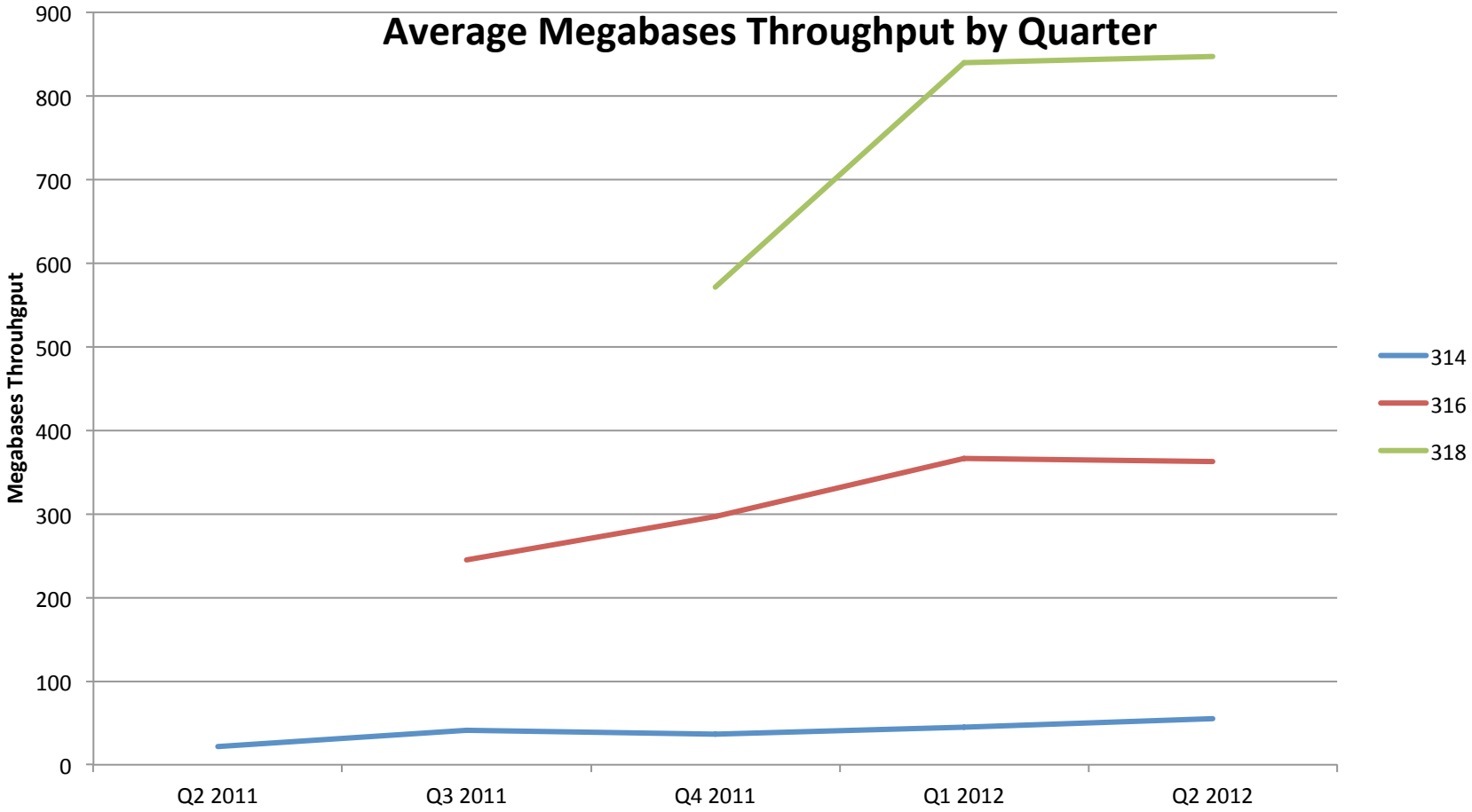
- Two Ion Torrent PGMs
  - Over 300 chips run to date
- Five SOLiD4 sequencing platforms
- One Life Techologies 5500XL
- One HiSeq 2000
- Automation thru Caliper Sciclone & Biomek FX
- Life Technologies Preferred Service Provider
- Agilent Certified Service Provider
- Commercial partnerships with companies such as CLCBio, DNANexus and Genologics
- MD/PhD & Masters Level Scientists and Bioinformaticians
- IT Infrastructure of >100 CPUs and >100TB storage

Agenda

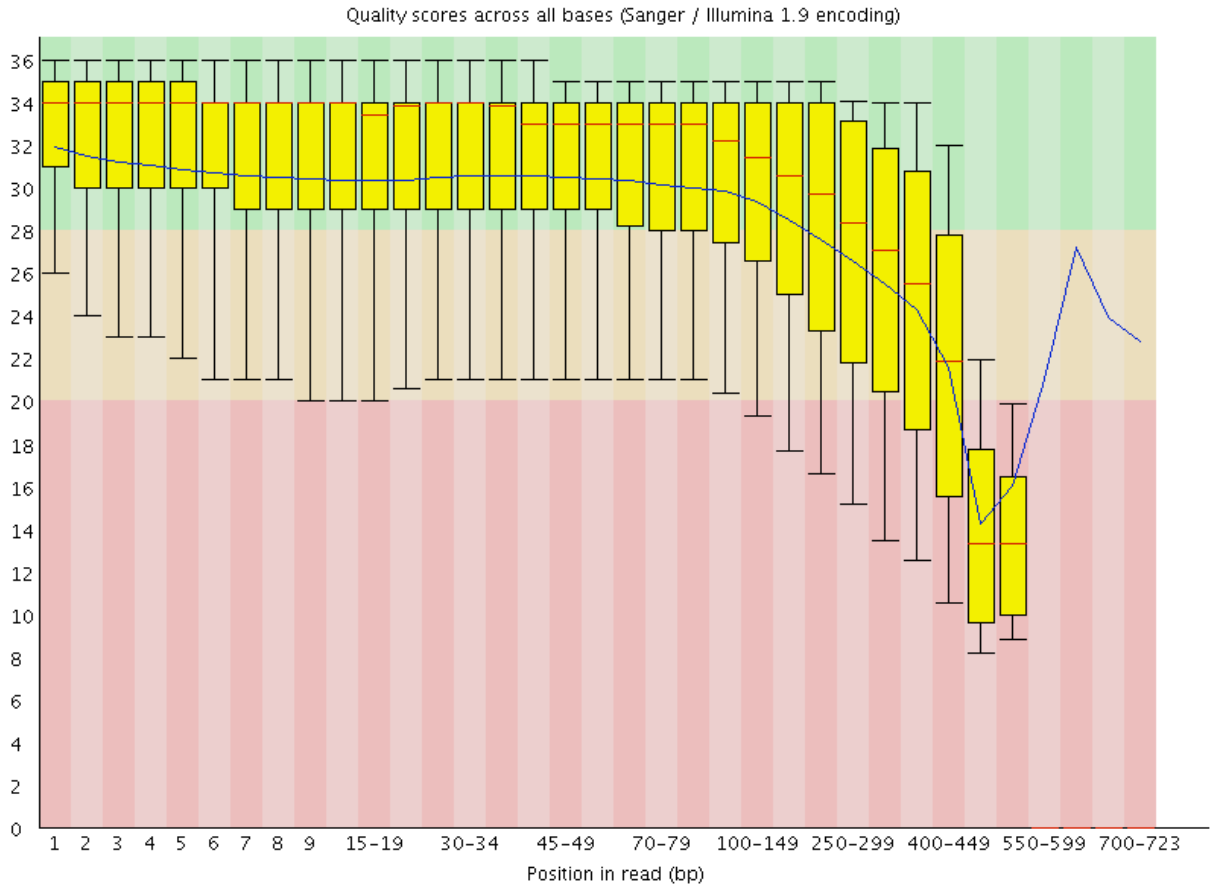
- Ion Torrent Sequencing
- Sequencing Applications
- Nimblegen Exome Capture
- Future of Semiconductor Sequencing

Ion Torrent Sequencing

Bigger Chips, Longer Reads



400bp Run Released Today



- No false positive INDEL calls
- Continued quality improvements could improve de novo assembly
- Majority of bases called >Q30

http://lifetech-it.hosted.jivesoftware.com/docs/DOC-2944

Improvements with TS 2.2

- Faster analysis times
- New GATK based variant caller
- Considers flow information for indel calling
- Baysian SNP caller
- Plugins from Community and LifeTech extend the Torrent Suite

Ion Torrent Applications

Ion Torrent Applications

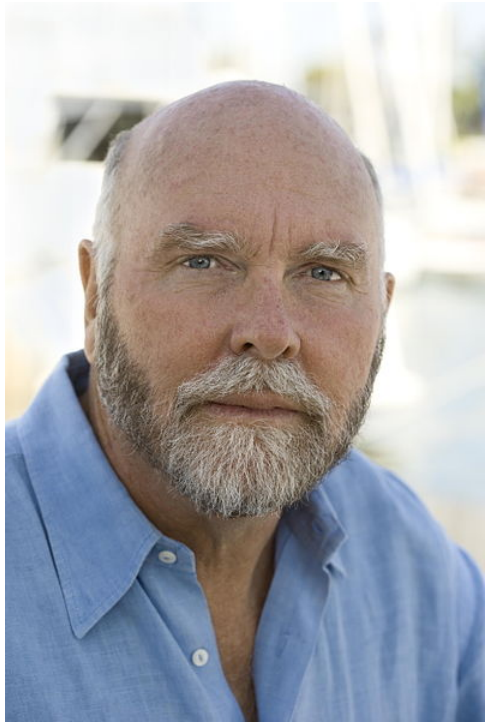
- de novo sequencing
- Resequencing
- Multiplexed Samples
  - 96 barcodes available
- Targeted Gene Panels
  - AmpliSeq multiplex PCR panel
    - AmpliSeq Cancer Panel
    - Comprehensive Cancer Panel
    - Inherited Disease Panel
    - Custom AmpliSeq Panels
- Exome Sequencing
  - Split over multiple chips

Nimblegen SeqCap Exome Capture

Nimblegen SeqCap Exome Library

- 64.1 Mbp Capture Kit
- HiSeq Exomes: ~50x coverage
- 318 Chip Current Throughput: 700 Mb
  - 5 Chips -> 50x Coverage with 100% on target reads
  - In reality on target not 100%
- 3x318 Chips
  - Proof of Concept
  - Pushing the limits of the PGM
  - Best case about 40x coverage
- Ion TargetSeq public exome runs released
  - 6 Chips at > 1GB each

Nimblegen SeqCap Exome Library



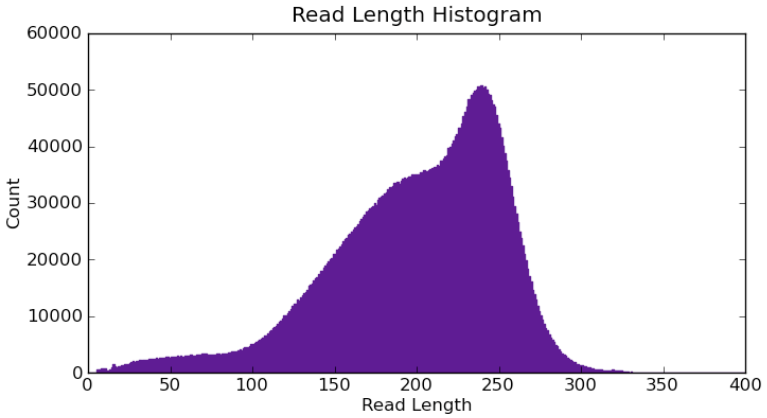
Goals

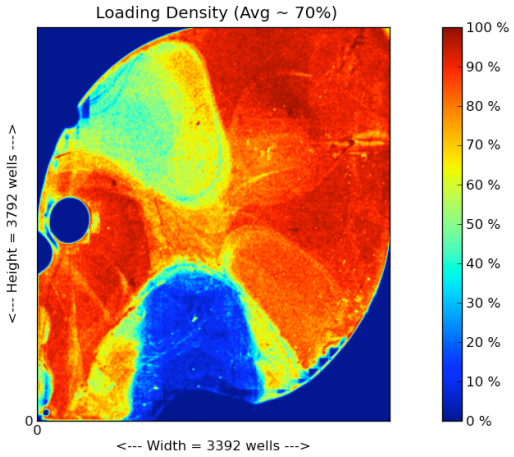
- Craig Venter HuRef DNA
- Prove feasibility of Nimblegen Exome Capture on PGM
- Show high concordance rates
- Show speed and automation from DNA to data

Limitations

- Comparatively expensive
- Low coverage

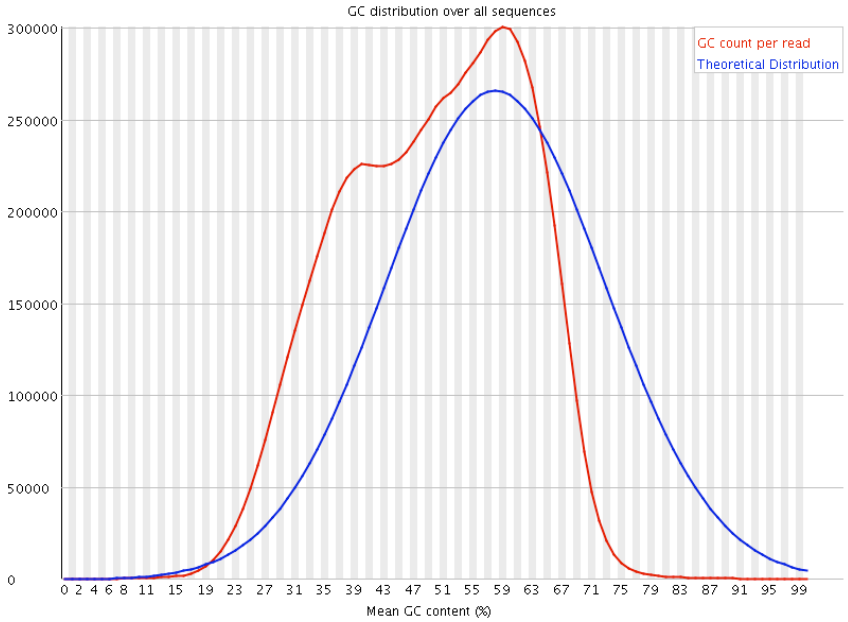
Results

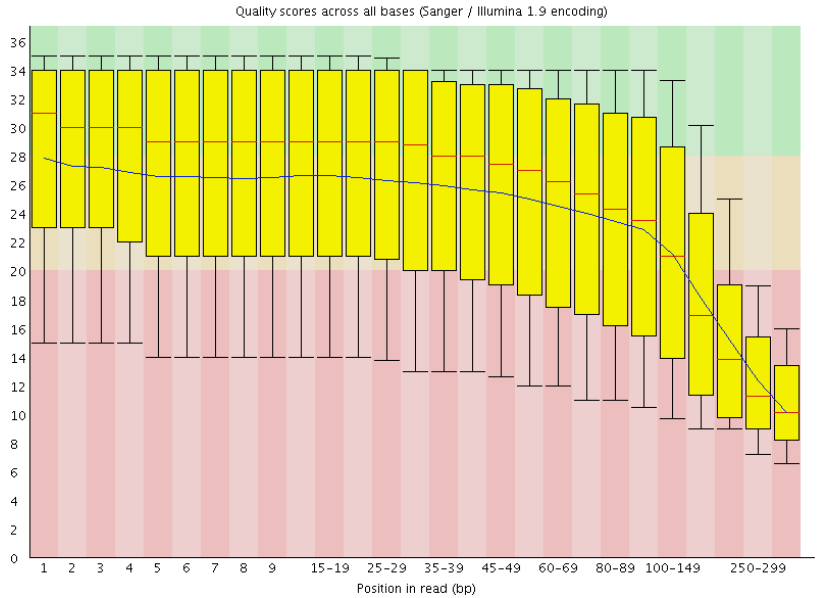




Library Summary	
Total Number of Bases (Mbp)	2,567.89
Q17 Bases (Mbp)	1,720.99
Q20 Bases (Mbp)	1,475.59
Number of Reads	12,840,960
Mean Length	182
Max Length	338

Results


GC distribution over all sequences
GC count per read
Theoretical Distribution
300000
250000
200000
150000
100000
50000
0
0 2 4 6 8 11 15 19 23 27 31 35 39 43 47 51 55 59 63 67 71 75 79 83 87 91 95 99
Mean GC content (%)


Quality scores across all bases (Sanger / Illumina 1.9 encoding)
36
34
32
30
28
26
24
22
20
18
16
14
12
10
8
6
4
2
0
1 2 3 4 5 6 7 8 9 15-19 25-29 35-39 45-49 60-69 80-89 100-149 250-299
Position in read (bp)

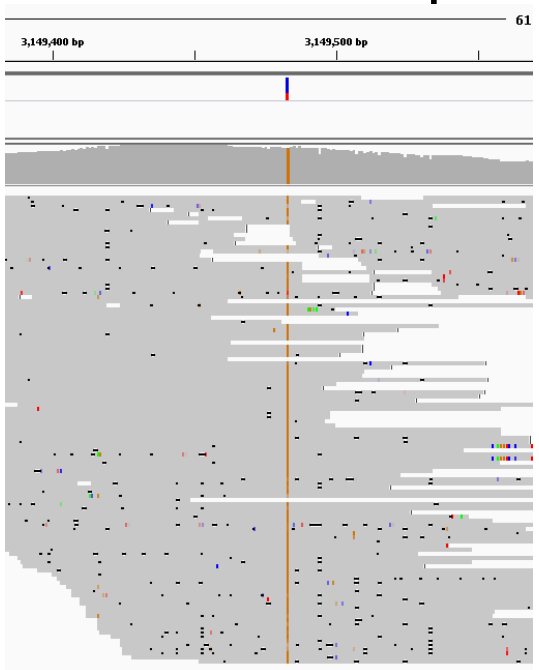
On Target Metrics

Number of reads on target	6,577,904
Percent reads on padded target	56.13%
Bases in targeted reference	64,190,747
Bases covered (at least 1x)	55,032,201
Average base coverage depth	13.42
Maximum base read depth	2,306
Average base read depth	15.67
Std.Dev base read depth	23.48
Target coverage at 1x	85.732%
Target coverage at 10x	41.598%
Target coverage at 20x	22.063%
Target coverage at 50x	4.359%
Target coverage at 100x	0.821%

Variant Calling

- 22,693 SNPs
  - 99.97% Concordance with HuRef
  - 75.95% Comparison rate
  - Heterozygous Calls: 9,328
  - Homozygous Calls: 13,365
  - Low Ti/Tv Ratio for novel SNPs – 0.77
- 46 INDELs
  - 81.82% Concordance with HuRef

- Low Coverage
- High concordance
- Quite a few false positives



Variant Annotation

- snpEff plugin
- Run after variantCaller finishes
- Report produced directly inside run report
- Available on Ion Community

Number of effects by impact

Type (alphabetical order)	Count	Percent
HIGH	102	0.61%
LOW	3,369	20.159%
MODERATE	2,677	16.018%
MODIFIER	10,564	63.212%

Number of effects by type and region

Type

Type (alphabetical order)	Count	Percent
DOWNSTREAM	1,331	7.964%
INTERGENIC	818	4.895%
INTRAGENIC	213	1.275%
INTRON	6,314	37.781%
NON_SYNONYMOUS_CODING	2,677	16.018%
NON_SYNONYMOUS_START	1	0.006%
SPLICE_SITE_ACCEPTOR	25	0.15%
SPLICE_SITE_DONOR	7	0.042%
START_GAINED	38	0.227%
START_LOST	3	0.018%
STOP_GAINED	55	0.329%
STOP_LOST	12	0.072%
SYNONYMOUS_CODING	3,327	19.908%
SYNONYMOUS_STOP	3	0.018%
UPSTREAM	859	5.14%
UTR_3_PRIME	788	4.715%
UTR_5_PRIME	241	1.442%

Region

Type (alphabetical order)	Count	Percent
DOWNSTREAM	1,331	7.964%
EXON	6,078	36.369%
INTERGENIC	818	4.895%
INTRON	6,314	37.781%
NONE	213	1.275%
SPLICE_SITE_ACCEPTOR	25	0.15%
SPLICE_SITE_DONOR	7	0.042%
UPSTREAM	859	5.14%
UTR_3_PRIME	788	4.715%
UTR_5_PRIME	279	1.669%

Compared with HiSeq

- TruSeq Capture Kit
  - CEU Daughter
  - 70.15% Reads on Padded Target
  - 37.44x Average Coverage Depth
  - Coverage at 20x: 73.071%
  - 99.84% Concordance with dbSNP
  - 98.60% Comparison rate
  - Heterozygous Calls: 26,299
  - Homozygous Calls: 17,239
  - Ti/Tv Ratio for all SNPs – 2.49
  - Ti/Tv Ratio for novel SNPs – 0.44
  - 94.7% INDEL Concordance
- Downsampled Nimblegen
  - Customer Sample
  - 90.76% Reads on Padded Target
  - 22.19x Average Coverage Depth
  - Coverage at 20x: 40.549%
  - 95.63% INDEL Concordance

Compared with TargetSeq on Ion

Number of reads on target	23,139,532
Percent reads on padded target	82.10%
Bases in targeted reference	37,268,825
Bases covered (at least 1x)	36,701,670
Average base coverage depth	79.87
Maximum base read depth	7,173
Average base read depth	81.11
Std.Dev base read depth	96.85
Target coverage at 1x	98.48%
Target coverage at 10x	95.88%
Target coverage at 20x	90.19%
Target coverage at 50x	59.02%
Target coverage at 100x	23.18%

- 22,333 SNPs
  - 99.80% Concordance with dbSNP
  - 97.71% Comparison rate
  - Heterozygous Calls: 12,304
  - Homozygous Calls: 10,029
  - Ti/Tv Ratio – All: 2.95
  - Ti/Tv Ratio – Known: 2.99
  - Ti/Tv Ratio – Novel: 1.69

http://lifetech-it.hosted.jivesoftware.com/docs/DOC-2659

Summary

	Ion Torrent	Ion Torrent	HiSeq	HiSeq
	Nimblegen	TargetSeq	Nimblegen (Down)	TruSeq
Number of Chips	3x318	6x318		
Read Lengths	1x200bp	1x200bp	2x100bp	2x100bp
Total Megabases	2,567.89	6,978.54	2,593.57	5,215.48
Total Reads	12,840,960	29,264,117	25,678,936	51,638,450
Sample	HuRef	NA12878	Customer	NA12878
Reads on Target	6,577,904	23,139,532	18,129,802	27,970,822
Percent on Padded Target	56.13%	82.10%	90.76%	70.15%
Average Depth	13.42	79.87	22.19	37.44
Target Coverage at 20x	22.06%	90.19%	40.55%	73.07%
SNPs	22,693	22,333	51,889	49,215
Concordance	99.97%	99.80%	99.91%	99.84%
Comparison Rate	75.95%	97.71%	97.83%	98.60%
Heterozygous Call	9,328	12,304	25,458	26,299
Homozygous Calls	13,365	10,029	18,832	17,239
Ti/Tv Total	2.05	2.95	2.58	2.49
Ti/Tv Novel	2.05	1.69	2.42	0.44

Future of Ion Torrent Sequencing

Proton

- 2 exomes per run at launch
  - >50x average coverage
- Similar data analysis time
- Same analysis pipeline as Ion Torrent
- No bioinformatics involvement until after Alignment/Variant Calling/ Annotation Steps

Gene Panels

- Targeted Sequencing
- Multiplex PCR means capture is very rapid
- Very high coverage
- Variant Caller can detect down to ~5% variant frequency
- 10ng DNA starting material

Conclusion

- Exome capture with high concordance possible with semiconductor sequencing
- Torrent Suite software allows automation of entire exome pipeline
- Higher coverage needed for more specific variant calling
- Ion Proton will provide sufficient coverage for 2 high coverage exomes per run

Thanks

Edge Bio Team

- Lab
  - Joy Adigun
  - Rossio Kersey
  - Jackie Jackson
  - Jennifer Sheffield
  - Ryan Mease
  - Aaron Johnson
- Informatics
  - Anju Varadarajan
  - Vani Rajan
  - Karthik Kota
  - Phil Dagosto
- Justin Johnson
- John Seed
- Dean Gaalaas

Follow Us:

- EdgeBio Twitter: @EdgeBio
- David Jenkins Twitter: @dfjenkins3
- Justin Johnson Twitter: @BioInfo
- djenkins@edgebio.com
- http://www.edgebio.com/blog/